ESTIMAÇÃO INTERVALAR

Inferência Estatística I

Professora: Renata S. Bueno

Introdução

- Ao invés de estimar o valor real de um parâmetro através de um número, poderíamos pensar em estimar um intervalo que pode conter este valor real.

- Neste caso, falamos em estimação intervalar.

- Por exemplo, estima-se que o número de desempregados em certa área seja 2,4 + ou – 0,3 milhões.

- Estamos quase certos de que o número real de desempregados esteja entre 2,1 e 2,7 milhões.

Intervalo de confiança

Definição: Seja $\boldsymbol{X} = (X_1, \ldots, X_n)$ uma amostra aleatória de uma variável com distribuição que depende de um parâmetro θ. Seja $g(\theta)$ uma função de θ. Considere $T_1 = t_1(X_1, \ldots, X_n)$ e $T_2 = t_2(X_1, \ldots, X_n)$ duas estatísticas satisfazendo $T_1 \leq T_2$ que possuem a seguinte propriedade para todos os valores de θ:

$$P(T_1 < g(\theta) < T_2) \geq \gamma.$$

O intervalo (T_1, T_2) é chamado de intervalo de confiança de γ para $g(\theta)$. Pode ser chamado também de intervalo de confiança de $\gamma 100\%$ para $g(\theta)$.

Se a inequação "$\geq \gamma$" for uma igualdade para todo θ, o intervalo de confiança é chamado exato.

Intervalo de confiança

- Depois dos valores das variáveis aleatórias $X_1, \ldots, X_n$ serem observados, os valores $T_1 = t_1$ e $T_2 = t_2$ são computados e o intervalo (t_1, t_2) é chamado de valor observado do intervalo de confiança.

Observação: Existe a possibilidade de que uma das duas estatísticas T_1 ou T_2 seja constante (somente uma delas). Em outras palavras, um dos dois limites do intervalo aleatório (T_1, T_2) poderia ser constante.

Intervalo de confiança

Definição: Intervalo de confiança unilateral.

Seja $X_1, \dots, X_n$ uma amostra aleatória da função de densidade (ou de probabilidade) $f(x|\theta)$. Considere a estatística $T_1 = t_1(X_1, \dots, X_n)$ tal que

$$P(T_1 < \tau(\theta)) \geq \gamma,$$

então T_1 é chamado de limite unilateral inferior de confiança para $\tau(\theta)$.

Similarmente, considere a estatística $T_2 = t_2(X_1, \dots, X_n)$ tal que

$$P(\tau(\theta) < T_2) \geq \gamma,$$

então T_2 é chamado de limite unilateral superior de confiança para $\tau(\theta)$.

Intervalo de confiança

Exemplo 11.1: Suponha que X_1, X_2, X_3, X_4 é uma amostra aleatória de 4 observações obtidas de uma população $N(\mu, 9)$. Considere: $X_1 = 1{,}2$; $X_2 = 3{,}4$; $X_3 = 0{,}6$; e $X_4 = 5{,}6$.

O EMV de μ é $\bar{X} = 2{,}7$. Queremos determinar os limites inferior e superior que formam um intervalo que pode conter o verdadeiro valor de μ.

Note que

$$Z = \frac{\bar{X} - \mu}{\sqrt{9/n}} = \frac{\bar{X} - \mu}{\sqrt{9/4}} = \frac{\bar{X} - \mu}{3/2} \sim N(0{,}1).$$

Intervalo de confiança

A função de densidade desta variável não depende do parâmetro μ. Portanto, podemos calcular a probabilidade de Z estar entre dois quaisquer número $l_1 < l_2$. Seja,

$$P(-1{,}96 < Z < 1{,}96) = 0{,}95,$$

Portanto, temos que

$$P\left(\bar{X} - 1{,}96\left(\frac{3}{2}\right) < \mu < \bar{X} + 1{,}96\left(\frac{3}{2}\right)\right) = 0{,}95.$$

Substituindo $\bar{X} = 2{,}7$ temos

$$P(-0{,}24 < \mu < 5{,}64) = 0{,}95.$$

Intervalo de confiança

O intervalo aleatório $\left[\bar{X} - 1{,}96\left(\frac{3}{2}\right); \bar{X} + 1{,}96\left(\frac{3}{2}\right)\right]$ e o resultado $[-0{,}24; 5{,}64]$ são chamados de intervalo de confiança de 95%. Cuidado!

Fique atento ao seguinte detalhe: O intervalo $[-0{,}24; 5{,}64]$ é uma observação obtida do intervalo aleatório $\left[\bar{X} - 1{,}96\left(\frac{3}{2}\right); \bar{X} + 1{,}96\left(\frac{3}{2}\right)\right]$ quando usamos o resultado amostral $\bar{X} = 2{,}7$.

Se a amostra mudar, será obtido outro valor para $\bar{X}$ e, consequentemente, o intervalo resultante será diferente de $[-0{,}24; 5{,}64]$.

Intervalo de confiança

Uma interpretação apropriada para $\left[\bar{X} - 1{,}96\left(\frac{3}{2}\right); \bar{X} + 1{,}96\left(\frac{3}{2}\right)\right]$ será: "a probabilidade de que neste intervalo aleatório contenha o valor real (desconhecido) de μ é de 0,95".

X_1	X_2	X_3	X_4	$\bar{X}$	$[\bar{X} - 1,96(3/2); \bar{X} + 1,96(3/2)]$
1,2	3,4	0,6	5,6	2,7	[-0,24 ; 5,64]
3,9	2,1	4,5	3,1	3,4	[0,46 ; 6,34]
-1,4	1,0	-2,1	-2,0	-1,1	[-4,04 ; 1,84]
⋮	⋮	⋮	⋮	⋮	⋮

Nesta tabela espera-se que aproximadamente 95% dos intervalos calculados contenham o valor verdadeiro de μ.

Intervalo de confiança

Tem-se, portanto, 95% de confiança de que o intervalo observado neste problema $[-0{,}24; 5{,}64]$ irá conter a verdadeira média μ.

A medida de nossa confiança é, neste caso, 0,95 (também chamada de coeficiente de confiança)

- Um intervalo com qualquer grau desejado de confiança pode ser obtido.

- Na verdade, é possível construir inúmeros intervalos com o mesmo coeficiente de confiança.

Intervalo de confiança

- Quaisquer "t_1" e "t_2", tal que a área abaixo da função de densidade entre t_1 e t_2, seja igual a 0,95, formam um intervalo de confiança de 95%.

- Iremos preferir aquele intervalo de confiança que for mais curto, isto é, escolhemos "t_1" e "t_2" o mais próximo possível um do outro de forma que a igualdade $P(t_1 < Z < t_2) = 0{,}95$ se mantenha válida.

- No caso da distribuição $N(0{,}1)$, dada sua simetria em relação a zero, a escolha de (t_1, t_2) que minimiza a amplitude $t_2 - t_1$ para uma área de probabilidade fixada será $t_2 = -t_1$.

Intervalo de confiança

Observação: Dado um intervalo de confiança de $100\,\gamma\%$ para θ, temos que um intervalo de confiança de $100\,\gamma\%$ para $\tau(\theta)$ pode ser obtido aplicando a função $\tau(\cdot)$.

Restrição: $\tau(\cdot)$ deve ser função estritamente monótona.

Exemplo 11.2: Se $\tau(\cdot)$ é uma função monótona crescente e (T_1, T_2) é um intervalo de confiança de $100\,\gamma\%$ para θ, então $(\tau(T_1), \tau(T_2))$ é um intervalo de confiança de $100\,\gamma\%$ para $\tau(\theta)$ visto que:

$$P[\tau(T_1) < \tau(\theta) < \tau(T_2)] = P[T_1 < \theta < T_2] = \gamma.$$

Método da quantidade pivotal

- Vamos descrever um método para encontrar intervalos de confiança.

- Assume-se uma amostra aleatória $X_1, \ldots, X_n$ de alguma função de densidade (ou de probabilidade) $f(x|\theta)$ indexada pelo parâmetro θ.

- O objetivo é encontrar um intervalo de confiança estimado para $\tau(\theta)$ que é uma função real de θ (θ pode ser um vetor de parâmetros).

Método da quantidade pivotal

Definição: Seja $X_1, \dots, X_n$ uma amostra aleatória de $f(x|\theta)$. Considere $Q = q(X_1, \dots, X_n, \theta)$, isto é, Q é uma função de $X_1, \dots, X_n$ e θ. Se Q tem distribuição que não depende de θ, então Q é chamado de quantidade pivotal.

Exemplo 11.3: Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição $N(\theta, 9)$. Qual seria uma quantidade pivotal neste caso?

- Usaremos a quantidade pivotal para obter um intervalo de confiança.

Método da quantidade pivotal

Método da quantidade pivotal:

Se $Q = q(X_1, \dots, X_n, \theta)$ é uma quantidade pivotal, então Q tem distribuição que não depende de θ. Para qualquer $0 < \gamma < 1$ fixo, existirá q_1 e q_2 dependendo de γ tal que

$$P(q_1 < Q < q_2) = \gamma.$$

Se para cada possível amostra $(x_1, \dots, x_n)$ temos:

$$q_1 < q(X_1, \dots, X_n, \theta) < q_2 \Leftrightarrow t_1(X_1, \dots, X_n) < \tau(\theta) < t_2(X_1, \dots, X_n),$$

sendo t_1 e t_2 funções que não dependem de θ, então $(t_1(X_1, \dots, X_n),\ t_2(X_1, \dots, X_n))$ é um intervalo de confiança de $100\,\gamma\,\%$ para $\tau(\theta)$.

Método da quantidade pivotal

Alguns comentários:

1. q_1 e q_2 não dependem de θ.

2. Para uma escolha de γ, existem muitos possíveis pares de números (q_1, q_2) que levam ao resultado $P[q_1 < Q < q_2] = \gamma$.

3. Diferentes pares (q_1, q_2) irão fornecer diferentes funções t_1 e t_2.

4. Deveríamos tentar selecionar aquele par (q_1, q_2) que faz t_1 e t_2 serem o mais próximo possível um do outro.

Método da quantidade pivotal

- A característica principal do método da quantidade pivotal é que a desigualdade $q_1 < q(X_1, \dots, X_n, \theta) < q_2$ pode ser reescrita como $t_1(X_1, \dots, X_n) < \tau(\theta) < t_2(X_1, \dots, X_n)$ para qualquer possível amostra $(x_1, \dots, x_n)$.

Exemplo 11.4: Seja $X_1, \dots, X_n$ uma amostra aleatória da $N(\theta, 1)$. Nosso objetivo é encontrar um estimador intervalar usando o método da quantidade pivotal para estimar $\tau(\theta) = \theta$.

Exemplo 11.5: Seja $X_1, \dots, X_n$ uma amostra aleatória da $N(\mu, \sigma^2)$. Encontre um intervalo de confiança para μ usando o método da quantidade pivotal quando σ^2 é conhecido e desconhecido.

Exemplos

Exemplo 11.6: Seja $X_1, \ldots, X_n$ uma amostra aleatória da $N(\mu, \sigma^2)$. Encontre um intervalo de confiança para σ^2 usando o método da quantidade pivotal supondo que μ é desconhecido.

Exemplo 11.7: Seja $X_1, \ldots, X_n$ uma amostra aleatória da Exp(θ). Encontre uma quantidade pivotal para θ e construa um intervalo de confiança para θ.

Exemplo 11.8: Seja $X_1, \ldots, X_n$ uma amostra aleatória da $U(0, \theta)$. Verifique se $\frac{X_{(n)}}{\theta}$ é uma quantidade pivotal para θ. Construa um intervalo de confiança para θ considerando que as probabilidades das caudas são iguais.

Duas amostras independentes

- Vamos considerar o caso em que temos $X_1, \dots, X_n$ uma amostra aleatória da variável aleatória $X \sim N(\mu_1, \sigma^2)$ e $Y_1, \dots, Y_m$, uma amostra aleatória da variável aleatória $Y \sim N(\mu_2, \sigma^2)$, onde X e Y são independentes.

- Suponha que existe interesse em estimar a quantidade $\theta = \mu_1 - \mu_2$.

- Verifique se $\bar{X} - \bar{Y}$ é um estimador não viciado para θ.

- Como poderíamos construir um intervalo de confiança para θ?

Duas amostras independentes

- Sabemos que

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right)$$

de modo que, sendo $\theta = \mu_1 - \mu_2$, consideramos a quantidade pivotal

$$Q(\boldsymbol{X}, \boldsymbol{Y}, \theta) = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0,1).$$

Duas amostras independentes

- Sendo σ^2 conhecido, temos, o intervalo:

$$\left[\bar{X} - \bar{Y} - z_{\frac{1+\gamma}{2}}\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}\ ;\ \bar{X} - \bar{Y} + z_{\frac{1+\gamma}{2}}\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}\right].$$

- Sendo σ^2 desconhecido, temos que uma quantidade pivotal é dada por:

$$Q(\boldsymbol{X}, \boldsymbol{Y}, \theta) = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

Duas amostras independentes

- Onde

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{(n+m-2)} \quad ; \quad S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\text{e} \quad S_y^2 = \frac{1}{m-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

- Como $\frac{(n-1)S_x^2}{\sigma^2} \sim \chi_{n-1}^2$ e $\frac{(m-1)S_y^2}{\sigma^2} \sim \chi_{m-1}^2$, pela independência de S_x^2 e S_y^2 temos que

$$\frac{(n+m-2)S_p^2}{\sigma^2} \sim \chi_{n+m-2}^2.$$

Duas amostras independentes

- Então, um intervalo de γ de confiança para θ é dado por:

$$\left[\bar{X} - \bar{Y} - t_{\frac{1+\gamma}{2};n+m-2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \;\; ; \;\; \bar{X} - \bar{Y} + t_{\frac{1+\gamma}{2};n+m-2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}}\right].$$

- Como ficaria o intervalo de confiança para σ^2?

Duas amostras independentes

- Suponha que temos $X_1, \ldots, X_n$ uma amostra aleatória de $X \sim N(\mu_1, \sigma_1^2)$ e $Y_1, \ldots, Y_m$, uma amostra aleatória de $Y \sim N(\mu_2, \sigma_2^2)$, onde X e Y são independentes, e o interesse é a construção de um intervalo de confiança para σ_1^2/σ_2^2.

- Note que

$$\frac{(n-1)S_x^2}{\sigma_1^2} \sim \chi_{n-1}^2 \quad \text{e} \quad \frac{(m-1)S_y^2}{\sigma_2^2} \sim \chi_{m-1}^2.$$

Duas amostras independentes

- Sendo assim, temos a seguinte quantidade pivotal:

$$Q(\boldsymbol{X}, \boldsymbol{Y}, \theta) = \frac{(m-1)S_y^2/\sigma_2^2(m-1)}{(n-1)S_x^2/\sigma_1^2(n-1)} \sim F_{m-1,n-1},$$

onde $F_{m-1,n-1}$ denota a distribuição F com $m-1$ e $n-1$ graus de liberdade.

- Então, dado γ, obtemos λ_1 e λ_2 na distribuição $F_{m-1,n-1}$, de modo que

$$P\left[\lambda_1 \leq \frac{\sigma_1^2 S_y^2}{\sigma_2^2 S_x^2} \leq \lambda_2\right] = \gamma.$$

Duas amostras independentes

- Considerando o intervalo com probabilidade iguais para as caudas, ou seja, $\lambda_1 = F_1$ e $\lambda_2 = F_2$, de modo que

$$P\left[F_{m-1,n-1} \geq F_2\right] = P\left[F_{m-1,n-1} \leq F_1\right] = (1-\gamma)/2,$$

onde F_1 e F_2 são obtidos na tabela da distribuição F com $m-1$ e $n-1$ graus de liberdade, temos o intervalo

$$\left[F_1 \frac{S_x^2}{S_y^2} \,;\, F_2 \frac{S_x^2}{S_y^2}\right].$$